

Access for all?

While initiatives for self-archiving and creating new open access journals gain momentum, new questions about the legal and economic basis of scientific publishing arise

Les Grivell

Last December, a UN-sponsored world summit on the information society (www.itu.int/wsis) and the announcement that the UK House of Commons Select Committee on Science & Technology intends to conduct an inquiry into access to scientific publications were two more developments in a movement that many regard as a crusade to free up the scientific literature. The basic aim—to provide toll-free access to the full text of published scientific articles for anyone wishing to read it—is clearly a noble one to which few would object. Nevertheless, in recent years the complexity of the issues and the difficulties involved in bringing about change in scientific publishing have become more apparent. For those who have not managed to keep up with the current complicated state of affairs, the following overview may provide food for thought.

The scientific world can be roughly divided into two groups: those with access to the scientific literature and those without. Members of the former group are often blissfully unaware of their privileged status. Their institutions have paid the required subscription fees, and for them the full text of journals is available at the click of a mouse. Those without access face the fact that virtually all recently published research is locked away behind toll barriers.

The movement to change this state of affairs arose out of what librarians call the 'serials crisis': the inability of institutional budgets to keep pace with the spiralling costs of scientific journals. The proponents of different solutions to the problem of access jostle shoulder to shoulder, each advocate aspiring to evangelical heights in an effort to promote their point of view. Broadly speaking, three mutually non-exclusive models specifically address the

issues of access. These are to start up new open access journals, to encourage authors to provide free access to self-maintained archives of previously published papers, and to open up access to conventional, subscription-based journals.

Note that in the previous paragraph I have used two terms, 'free' and 'open' access. This is not sloppiness on my part. These terms are in fact different and it is the nature of this distinction that is the subject of heated debate, not least because the terms have in the past often been used interchangeably, although they do have different connotations.

According to the 2002 Budapest Open Access Declaration (www.soros.org/openaccess), "open access" is defined as "free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself." The declaration goes on to define the rights of authors: "The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited."

In contrast, the 2003 Bethesda Declaration on Open Access defines open access more widely. Works have to meet two conditions: "The author(s) and copyright holder(s) grant(s) to all users a free, irrevocable, worldwide, perpetual right of access to, and a license to copy, use, distribute, transmit and display the work publicly and to make and

distribute derivative works, in any digital medium for any responsible purpose [...] A complete version of the work and all supplemental materials, including a copy of the permission as stated above, in a suitable standard electronic format is deposited immediately upon initial publication in at least one online repository that is supported by an academic institution, scholarly society, government agency, or other well-established organization [...]."

The proponents of different solutions to the problem of access jostle shoulder to shoulder, each advocate aspiring to evangelical heights in an effort to promote their point of view

In other words, under the Budapest concept of open access, an author can choose to make his work freely available, but retains the right to restrict the ability of others to copy or redistribute it as a whole, or in part. Under the Bethesda concept, the author waives his or her rights under copyright law. Publication in a Bethesda declaration-compliant journal is thus the only option for achieving full open access. This difference has fuelled recent vigorous mailing-list exchanges between Stevan Harnad from the University of Southampton, UK, and Michael Eisen from the University of California at Berkeley (Berkeley, CA, USA), as supporters of the Budapest and Bethesda definitions, respectively. Eisen redefines the Budapest concept as one that allows toll-free, but not open access. According to Eisen, in addition to publisher-implemented toll barriers, the lack of the right to reformat and/or redistribute the text is an unnecessary compromise of the interests of the scientific

community in situations in which full texts are used for computational or data-mining analysis. According to Harnad, indexing and other forms of text analysis can be performed under the Budapest declaration without asking authors to sign away their rights. In his opinion, this emphasis on new open access journals and copyright issues places unnecessary limits on the speed with which information in the literature can be made freely accessible.

Copyright restrictions, or rather their abolition, are also the central theme of a bill introduced in the US House of Representatives in June 2003 by Democratic congressman Martin Sabo from Minnesota (Sabo, 2003). The Public Access to Science Act aims to amend existing US copyright law to ensure that research "substantially funded" by the US government cannot be copyrighted and thus remains freely available to the public. Whether it will achieve this aim is hotly disputed.

Critics of the proposal have pointed out that the bill in fact does little to facilitate open access and many have also questioned the wisdom of removing copyright protection. Notable opponents include the Federation of American Societies for Experimental Biology (FASEB), the Association of American Universities (AAU) and many scientific society and commercial publishers. FASEB argues that the bill "threatens to destroy the current field of scientific publishing and will harm scientific societies that rely on publishing revenues to support other professional activities" (Wells, 2003). The AAU is concerned about long-term access to scholarly research, yet in an open letter to congressman Sabo, AAU president Nils Hasselmo stressed that copyright protection is an important means of ensuring the accuracy and authenticity of publications and that its removal is unnecessary for promoting public access (Hasselmo, 2003). In addition, removal of protection under the bill would adversely affect other copyrighted works including computer software, and thus reduce incentives for universities and industry to collaborate in technology transfer. In a vigorously worded editorial, Michael Held, Director of Rockefeller University Press, warned that lack of copyright protection would lead to pirating and uncontrolled reprinting of published materials, thereby strongly reducing the incentive of the original publisher to invest resources in their production (Held, 2003).



In an analysis of the Sabo Bill and the initial negative reactions to it, Samuel E. Trosow, Assistant Professor of Law at the University of Western Ontario in London, Canada, argued that some of these reactions are based on a misunderstanding of the purpose and scope of US copyright law and patentability under US Patent law (Trosow, 2003). Singling out the AAU for criticism, he remarked that "one is left with the distinct impression that the large research universities appear to view the Sabo Bill as a threat to their overall programme of commercialization, and that this consideration drives their opposition more so than any supposed deleterious affect the bill would have on the production and dissemination of works derived from STM [scientific, technical and medical] research."

Amid the hubbub of protests and actions from organizations likely to be affected by the Sabo Bill, one major player seems to have largely been ignored: the research scientist. In the role of reader, a scientist may vociferously demand free or open access. As authors, however, most scientists are concerned that their papers should be published in good journals and that they should be subsequently read and cited frequently. Although it has been claimed that publication in open access journals increases the chances of citation (Lawrence S, 2003), few open access journals in the life sciences have yet made it into the elite category. Younger scientists in particular are cautious about putting their careers at risk by submitting work to what they regard as unknown or sub-optimal journals. Even fewer may be willing to risk their careers if they either lose

control or fear loss of control over their works through legislation that removes copyright protection.

These fears are not without foundation. Under Canadian and European copyright, an author can transfer distribution rights to a publisher as part of the process of publication, but still retain his or her legally enforceable moral rights to attribution, integrity and association. Attribution allows an author to be associated by name with the published work. Integrity protects against "distortion, mutilation, or other modification". Association allows an author to control his or her work in association with a product, service, cause or institution. All three rights provide valuable safeguards in a situation where, for example, a company might seek to promote a product by selectively quoting from a published paper. Curiously, the moral rights of authors do not exist under US copyright legislation (Trosow, 2003). Thus the Sabo Bill cannot erode them further. I leave it to readers to draw their own conclusions as to whether they consider this good or bad news.

Academic as the above considerations may appear to someone who, as yet, has no access at all to the scientific literature, they do have important effects for both the speed and the mechanism by which access is achieved. Ignoring for a moment the fact that many profit and not-for-profit publishers make publications older than 12 months available free of charge, one notable feature of 2003 was a series of declarations affirming the intention of their signatories to support the establishment of new open access journals that meet the two conditions laid down by the Bethesda declaration. The USA-based Public Library of Science (www.plos.org) launched its high profile open access *Biology* journal and considerable publicity was devoted to promoting the view that this form of open access publication should be the way forward and would give scientists access to the literature, and freedom to create and distribute derivative works and to carry out sophisticated computational analysis or data-mining on reprocessed texts.

Others, such as Harnad, are more cautious. While encouraging publishers of both new open access and conventional scientific journals to make their content freely available, Harnad is a long-standing proponent of self-archiving. In this scheme, individual scientists or their institutions set up web servers

on which copies of previously published works are stored and made accessible by the use of protocols that allow search engines to harvest relevant bibliographic information or 'metadata'. He points out that open access journals currently represent only about 5% of the estimated 24,000 journals responsible for publishing a total of 2.5 million scientific articles per year. In his opinion, the fact that many publishers of conventional journals already permit self-archiving by authors means that free access to a very significant proportion of the world's scientific literature could in theory be achieved virtually overnight. Open access journals, in contrast, depend for their existence on largely untested models of economic operation, and, as for any other journal, take time to establish themselves and build up a scientific reputation.

In the role of reader, a scientist may vociferously demand free or open access. As authors, however, most scientists are concerned that their papers should be published in good journals and that they should be subsequently read and cited frequently

So, how does self-archiving work? Is it an effective mechanism to provide free access? If so, why has it not yet become established? Answering these questions in turn, the first step in self-archiving is to set up a web server with software that allows data to be stored, searched and retrieved. EPrints (www.eprints.org) and DSpace (www.dspace.org) are perhaps the best-known self-archiving software packages. They are open source, freely available and comply with the Open Archive Initiative (OAI). This means that the bibliographic metadata of the stored articles can in principle be harvested and shared among different self-archive servers, thereby allowing these to form interoperable nodes in a Napster-like network. Setting up and maintaining a server is a relatively simple task. Self-archiving is thus an effective and low-cost mechanism to provide free access. OAI-compliant search engines, such as the University of Michigan's OAIster, which covers approximately 2.4 million records from about 250 institutions, provide the means for finding and retrieving documents.

So why has a large part of the scientific literature not become freely accessible

overnight? As discussed below, there may be difficulties with some technical issues but the main problems seem to be social and psychological factors. A scan through the comments of, for example, the American Scientist Open Access Forum, suggests widespread ignorance of the potential of the approach and of the right of authors to self-archive; confusion between preprint and publication servers; lack of motivation on the part of individuals and/or sluggishness on the part of their institutions to adopt and finance a self-archiving policy.

The technical issues relate to searchability. More than in other literature retrieval systems, the quality of the metadata that characterize individual articles in a distributed self-archiving network is crucial to the success of cross-searching efforts. Unfortunately, institutional archives often need to cover a broad spectrum of scientific disciplines, so that the classification of articles is often limited to major subject classifiers only. Furthermore, the simple OAI-compliant metadata models (www.openarchives.org/OAI/openarchivesprotocol.html) used by most self-archiving software may introduce additional limitations on search flexibility. Metadata is often incomplete, and archive curators have to add to limited metadata manually. Clearly, only when a researcher is able to search rapidly and accurately through full-text articles will self-archiving become a more attractive proposition. Information platforms, such as E-BioSci (www.e-biosci.org), are likely to have an important role in realizing the full potential of self-archive repositories.

Given the primary aim of providing toll-free access to as much of the scientific literature as possible and three non-exclusive ways of achieving this, what are the likely consequences of each for the longer-term future of publishing? Let us examine each of the routes separately.

At first sight, self-archiving seems to be the least disruptive of the three routes and the most friendly to the world of conventional publishing. However, it takes little imagination to see that in the longer term a significant degree of self-archiving, coupled with efficient systems for search, retrieval and caching could have a serious impact on demand for subscription-based content, with a consequent decline in incentives for publishers to invest in these activities.

Although costs vary widely from journal to journal, there is no doubt that publishing is an expensive business in general, and open access journals are no different from other journals. A new journal can take between three and five years to establish itself and requires considerable capital investment during this period. The staffing costs associated with peer review and editing accepted manuscripts will depend on the volume of manuscripts and acceptance rates, but are significant. Online publications, although clearly able to reach more readers at a lower cost than paper-based journals, still need continued investment in technology and hardware. BioMed Central, a pioneering publisher in this field, with more than 100 journals under its wing, charges authors US\$525 for processing each manuscript. The Public Library of Science's elite *Biology* journal charges US\$1,500 per manuscript for processing. However, other publishers estimate that as about 90% of manuscripts are rejected to maintain quality, the real costs for this and similar journals may be anything up to an order of magnitude higher (Butler, 2003).

Publication is part of the research process, costs are associated with any publication system, open access or not...

Nevertheless several major funding institutions, including the Howard Hughes Medical Institute and the National Institutes of Health in the USA, the Deutsche Forschungsgemeinschaft and the Wellcome Trust in Europe, have so far been supportive of the principles of free or open access. They have pledged funds to meet the costs of institutional subscriptions to the journals or processing charges for individual manuscripts. How this will work out in the long term remains to be seen. Open access journals have not been in existence long enough for proper testing of this economic model. However, there could be painful surprises for laboratories with high publication levels, which could easily run up bills of many tens of thousands of Euros in this way.

Not surprisingly, few conventional publishers have so far rushed to convert existing subscription-based, access-limited journals to full open access. Some publishers—such as the Company of Biologists, the American Physiological Society and Oxford University Press—have established curious, hybrid models in which subscriptions are

maintained, but authors pay for open access to their own articles. Nevertheless, the *British Medical Journal's* early bold experiment in offering free access to all content has not been widely taken up. Worse, falling revenues from subscriptions have forced them to return to paid access. After 10 years of free access, a large part of BMJ's content will, from 2005 on, disappear behind access controls for "a year or more (after publication)" (Delamothe & Smith, 2003). Even though only 12% of the journal's income is generated from subscriptions with the rest coming from advertising, a 9% reduction in subscriptions in the past year coupled with other losses has led the BMJ Publishing Group Board to take this step, albeit reluctantly, to "defray costs of the journal's web-site and to allow funding of further developments."

The BMJ's difficulties should serve as a serious warning to those who have been calling for the abolition of the subscription-based system at any cost and with it much of the value added to scientific publications by publishers, whether for profit or not. Some proponents of the free or open access revolution seem to accept calmly the 'downsizing' or even termination of current journal publishing activities as inevitable but this would have dramatic consequences with the permanence, integrity, accuracy and independence of the scientific record at stake.

Although commercial publishers have been the subjects of the most virulent attacks so far, it is the learned societies perhaps who stand to lose most—the ability to support the very communities that founded them. This paradox is simplistically brushed aside in the FAQ of the self-archiving community: "Learned Societies are potential allies in and beneficiaries of the self-archiving initiative. First, they are us. Whatever is good for research, and for research impact, is therefore good for Learned Societies. [...] But many of them are also journal publishers, and hence may be facing downsizing pains. Unlike commercial publishers, however, their first and last allegiance will of course be to research and researchers, that is, us. We will hear rationalizations about needing the toll revenues to fund 'good works' such as meetings, scholarships and lobbying. But it will quickly become evident that some of these good works are not actually essential, and certainly nothing for which we would want to sacrifice research impact; and the subset of them that really is essential (such as meetings) will prove to be able to fund itself

in other ways, rather than needing to be subsidized at the expense of research impact" (www.eprints.org/self-faq).

In summary, the movement to achieve free and/or open access to the literature has put the world of scientific publishing into a state of flux that is leading to a re-examination of all aspects of the publishing process and the roles of publishers and scientists as producers, reviewers and consumers of the literature. As discussed above, the debate has gone beyond the simple question of access to the basic economics of publishing. Simple proposals for redistributing costs have raised complex questions about the ownership, authenticity and integrity of the scientific record itself.

At the same time, the increasing pressure from research institutions and funding bodies on authors to publish in new low- or no-cost journals has triggered heated discussion on the quality of both current and future publications, the value of peer review and different publication strategies (Lawrence PA, 2003), and about how scientists in their role as reviewers and advisors evaluate publications as a major component of the reward system in science (Colquhoun, 2003).

To the lay public, the complexity of the debate must be confusing in the extreme. The Public Library of Science's publicity campaigns in the USA have targeted the taxpayer. They point out that "the current closed system of publication places the narrow interests of publishers before the public interest and greatly diminishes the value of the more than US\$50 billion invested by US taxpayers each year in scientific and medical research." The implication that taxpayers pay twice for something that in the end they have no access to is clearly calculated to make anyone hot under the collar. But the full facts are much more complicated and difficult to grasp and are unlikely to lead to a change in public opinion. Publication is part of the research process, costs are associated with any publication system, open access or not, and the 'author-pays' open access models will simply lead to the reallocation of library funds to research budgets. This state of affairs is unlikely to change, as Margaret Reich, Director of Publications for the American Physiological Society, commented humorously: "[...] why should any of us, scientist and patient alike, have to pay again to read the results of that [federally funded] research? That sounds good, but some of my tax dollars also go to wheat and other farm subsidies,

and I don't see anyone handing me free loaves of Wonder Bread™" (Reich, 2003).

In the long term, no doubt access will be available for all, but at a price. Good (business) sense will hopefully prevail and hopefully too, the long-term benefits will outweigh the costs for all involved. Although likely to be just as ruthless in its progress, it probably will be evolution, rather than revolution (Owens, 2003) that will determine the future course of events. Changes in attitudes and economic models will lead to the creation of new niches from which both society and commercial publishers will be able to continue to serve the scientific community.

REFERENCES

- Butler D (2003) Scientific publishing: who will pay for open access? *Nature* **425**: 554–555
- Colquhoun D (2003) Challenging the tyranny of impact factors. *Nature* **423**: 479
- Delamothe T, Smith R (2003) Paying for bmj.com. *BMJ* **327**: 241–242
- Hasselmo N (2003) Letter from the AAU to M Sabo. www.aau.edu/intellect/sabo7.18.03.pdf
- Held MJ (2003) Proposed legislation supports an untested publishing model. *J Cell Biol* **162**: 171–172
- Lawrence PA (2003) The politics of publication. *Nature* **422**: 259–261
- Lawrence S (2003) Free online availability substantially increases a paper's impact. www.nature.com/nature/debates/e-access/articles/lawrence.html
- Owens SR (2003) Revolution or evolution? *EMBO Rep* **4**: 741–743
- Reich M (2003) Peace, love and PLoS. *Physiologist* **46**: 137
- Sabo M (2003) Public access to science act. H.R. 2613, US House of Representatives, 108th Congress, 1st session.
- Trosow SE (2003) Copyright protection for federally funded research: necessary incentive or double subsidy? http://publish.uwo.ca/~strosow/sabo_bill_paper.pdf
- Wells RD (2003) Letter from the Federation of American Societies for Experimental Biology to M Sabo. www.faseb.org/opa/news/docs/sabo.pdf



Les Grivell is manager of E-BioSci, the European electronic publishing initiative.
E-mail: les.grivell@embo.org

doi:10.1038/sj.embor.7400107